



Universiteit
Leiden
The Netherlands

Permutation-based inference for high-dimensional data.

Hemerik, J.

Citation

Hemerik, J. (2018, May 1). *Permutation-based inference for high-dimensional data*. Retrieved from <https://hdl.handle.net/1887/61825>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/61825>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/61825> holds various files of this Leiden University dissertation.

Author: Hemerik, J.

Title: Permutation-based inference for high-dimensional data

Issue Date: 2018-05-01

Permutation-based inference for high-dimensional data

Jesse Hemerik

Printing: Off Page, the Netherlands

©2018 Jesse Hemerik, Leiden, the Netherlands.

All rights reserved. No part of this publication may be reproduced without prior permission of the author.

ISBN: 978-94-6182-885-9

Permutation-based inference for high-dimensional data

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 1 mei 2018
klokke 16.15 uur

door

Jesse Hemerik
geboren te Leiden in 1989

Promotores:

Prof. dr. J. J. Goeman

Dr. A. Solari

· *University of Milano-Bicocca, Milaan*

Leden promotiecommissie:

Prof. dr. S. le Cessie

Prof. dr. A. W. van der Vaart

· *Universiteit Leiden en Universiteit van Amsterdam*

Prof. dr. M. A. van de Wiel

· *Vrije Universiteit Medisch Centrum en Vrije Universiteit,
Amsterdam*

Table of Contents

1	Introduction	1
1.1	High-dimensional data: examples	1
1.2	Multiple testing and the false discovery proportion	3
1.3	Dependence and permutation methods	4
1.4	This thesis	6
1.4.1	Permutation-based confidence on the false discovery proportion	6
1.4.2	Robust testing in generalized linear models	8
2	Exact testing with random permutations	9
2.1	Introduction	10
2.2	Fixed transformations	11
2.2.1	Basic permutation test	11
2.2.2	Permutation p -values	15
2.3	Random transformations	16
2.3.1	Comparison of Monte Carlo and permutation tests	17
2.3.2	Estimated p -values	18
2.3.3	Random permutation tests	18
2.3.4	p -values based on random transformations	23
2.4	Applications	24
2.5	Discussion	25
3	False discovery proportion estimation by permutations: confidence for SAM	27
3.1	Introduction	28
3.2	Basic upper bound	30
3.2.1	Setting and notation	30
3.2.2	Upper bound and median unbiased estimate	32
3.2.3	Choice of rejection regions	34

3.3	Closed testing for improved bounds	35
3.3.1	Closed testing	35
3.3.2	Improved FDP bounds	36
3.3.3	Approximation method	38
3.4	Conservative shortcut	39
3.5	Simulations	43
3.5.1	Simulated data and tests used	43
3.5.2	Performance of variants of SAM as bounds	44
3.5.3	Performance of variants of SAM as estimators	45
3.5.4	Performance of the closed testing-based bound	47
3.5.5	Performance of the conservative shortcut	47
3.5.6	Performance of the approximation method	48
3.6	Application to data	50
3.7	Discussion	53
4	Permutation-based simultaneous confidence bounds for the false discovery proportion	55
4.1	Introduction	56
4.2	Setting and notation	58
4.3	Single-step procedures	59
4.3.1	Confidence envelopes	59
4.3.2	Parametric confidence envelopes	60
4.3.3	Meinshausen's nonparametric confidence envelope	61
4.3.4	Examples of candidate envelopes	63
4.4	Improved bounds by closed testing	65
4.5	Iterative method	67
4.5.1	Exact method	67
4.5.2	Approximation method	68
4.6	Simulations	69
4.6.1	Performance of the iterative method	70
4.6.2	Performance of the approximation method	70
4.7	Data analysis	71
4.8	Discussion	75
5	Robust testing in generalized linear models by sign-flipping score contributions	77
5.1	Introduction	77
5.2	Models with known nuisance parameters	79

5.2.1	Basic sign-flipping test	80
5.2.2	Fixed transformations	82
5.2.3	Asymptotic equivalence with parametric test	83
5.2.4	Robustness	84
5.3	Taking into account nuisance estimation	86
5.3.1	Asymptotically exact test	87
5.3.2	Robustness	90
5.3.3	An example	91
5.4	Simulations	92
5.4.1	Overdispersion, heteroscedasticity and estimated nuisance	93
5.4.2	Ignored nuisance	94
5.4.3	Power	95
5.4.4	Strong heteroscedasticity	96
5.5	Data analysis	97
5.6	Discussion	98
A Manual of the R package <i>confSAM</i>		101
Bibliography		109
List of Publications		115
Samenvatting		117
Dankwoord		123
Curriculum Vitae		125

